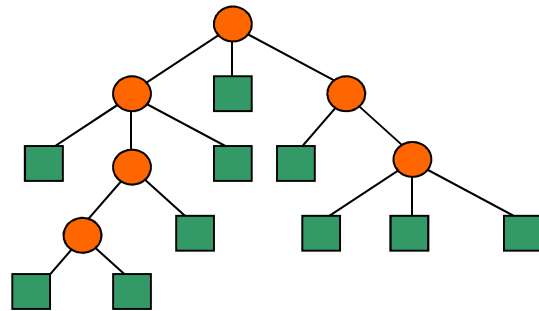# Beam Search Extraction and Forgetting Strategies on Shared Ensembles

Vicent Estruch, Cèsar Ferri,
José Hernández-Orallo, M.José Ramírez-Quintana
`{vestruch, cferri, jorallo, mramirez}@dsic.upv.es`
Dep. de Sistemes Informàtics i Computació,
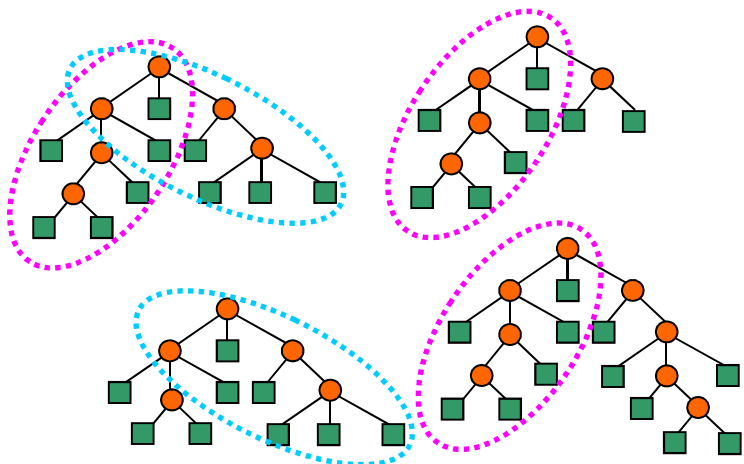*Universitat Politècnica de València,*
Valencia, Spain

Ensembles of Decision Trees

- ## Decision Tree:

  - Each internal node represents a condition.
  - Each leaf assigns a class to the examples that fall under that leaf.

- ## Forest: several decision trees can be constructed.

  - Many trees have common parts.

  - Traditional ensemble methods repeat those parts:
    - memory and time ↑↑↑.
    - comprehensibility is lost.

Decision Tree *Shared* Ensembles

- Shared ensemble:
  - Common parts are shared in an AND/OR tree structure.

- Construction space and time resources are highly reduced.

- Throughput is also improved by this technique.
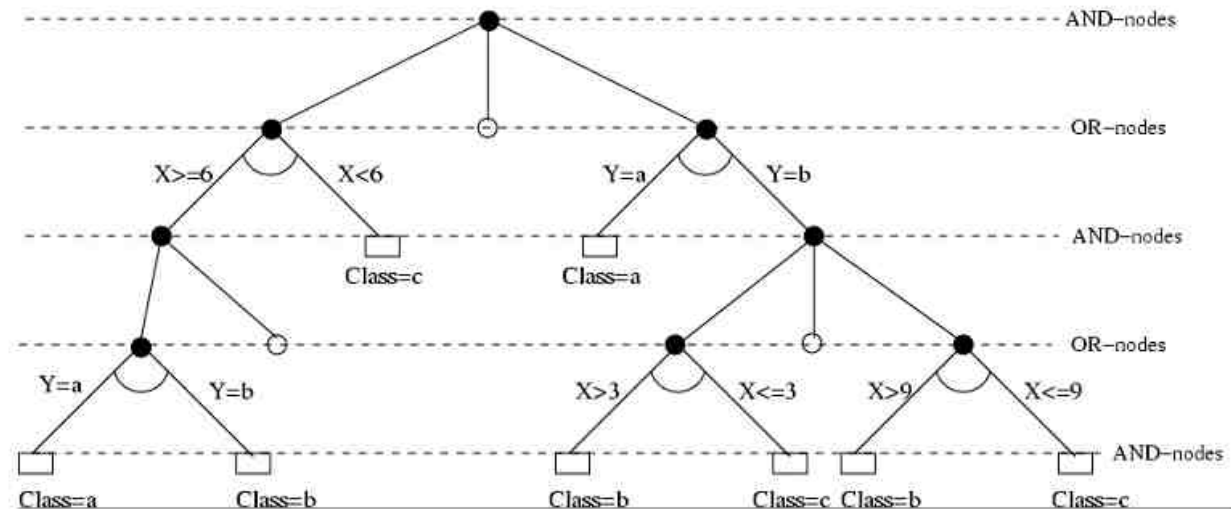
Multi-tree Construction

- Suspended OR-node selection criteria:
  - Rival ratio
  - Rival absolute
  - Random
  - Topmost
  - Bottom

Node Forgetting

- Method to filter some of the suspended nodes.

- Reduction of the computational resources

- Studied forgetting methods:
  - Constant
  - Logarithmic
  - Logarithmic + Depth

Experiments (1/4)

- Experimental setting:

  - 23 datasets from the UCI repository.

  - 10 X 10 Cross Validation.

  - Multi-tree implemented in the **SMILES** system.

  - Splitting criterion: GainRatio (C4.5).

  - Boosting and Bagging from WEKA.

Experiments (2/4)

- Comparison among construction criteria:

28 datasets	Bottom	Rival Abs	Random	Rival Rat	Topmost
Acc. GeoMean	80.56	80.70	82.45	80.69	83.61
Sec. GeoMean	0.10	0.14	1.34	0.46	9.12

- Comparison among forgetting criteria:

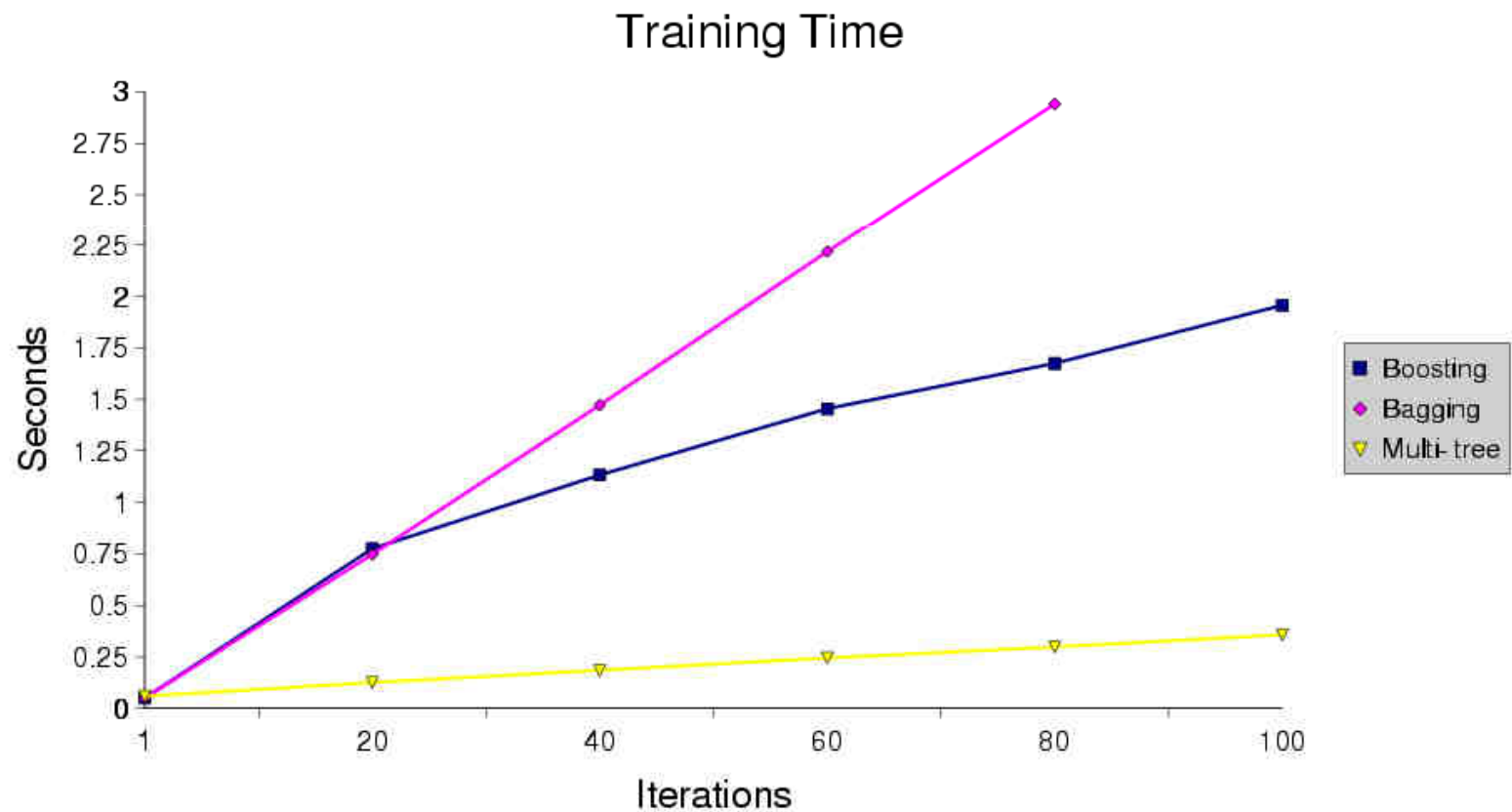| 28 datasets	No forg	Const=5	Log.	Log+Dep.
Acc. GeoMean	82.45	82.22	82.38	82.55
Sec. GeoMean	1.34	1.16	0.99	1.23

Experiments (3/4)

- Combination Accuracy compared to other Ensemble Methods:
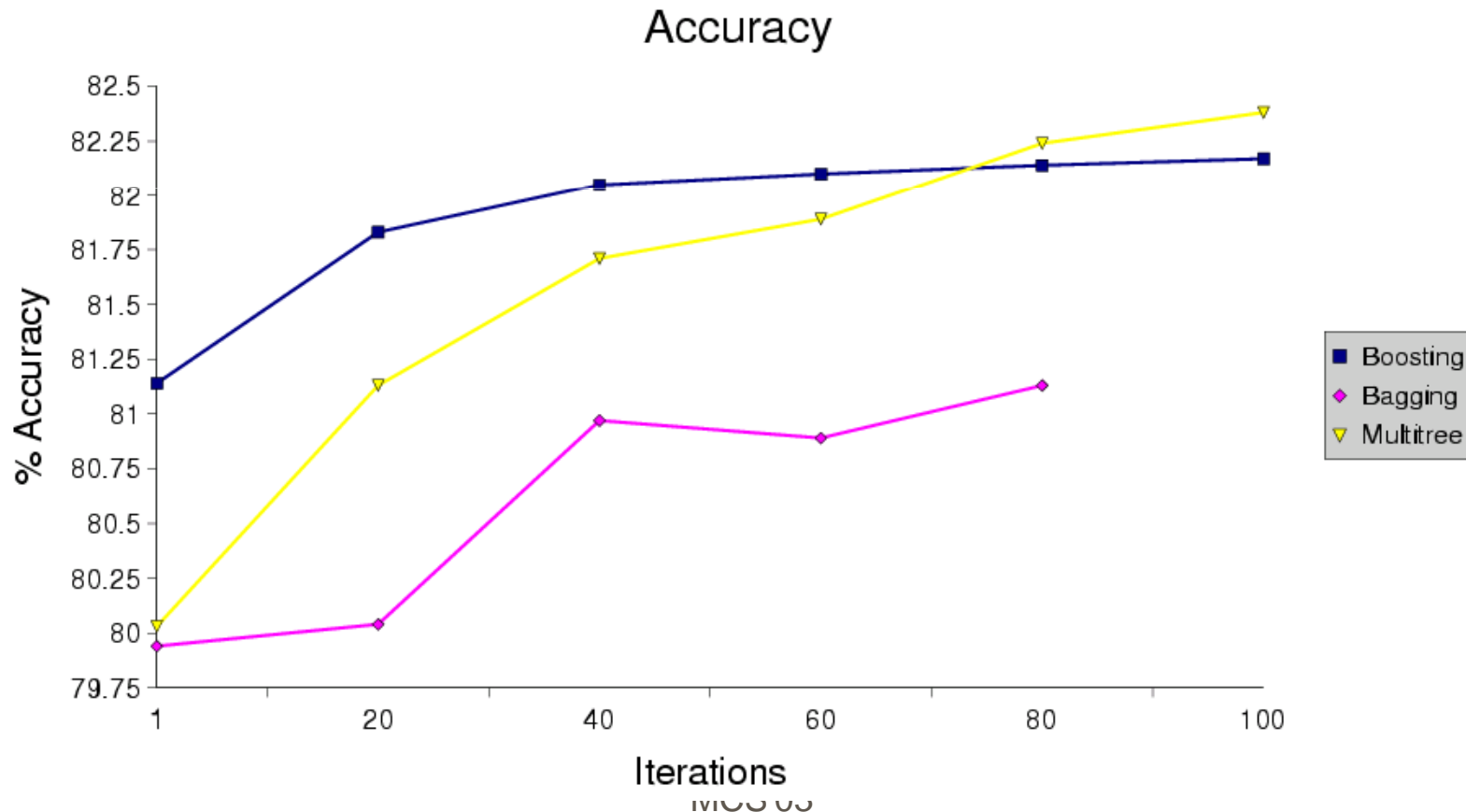
Experiments (4/4)

- Combination Resources compared to other Ensemble Methods:

Training Time

Conclusions

- Multi-tree as an alternative to other population strategies for shared decision tree ensembles:
  - Anytime character
    - The first tree is obtained in the same way as classical eager decision tree learning.
    - We ask for further solutions on demand.
  - Population (and hence resources) is scalable and easy to be controlled.
- Same or even better accuracy results than other ensemble methods with significantly lower resource consumption.